The Sapphire Project: Scientific Data Analysis

Chandrika Kamath

Center for Applied Scientific Computing

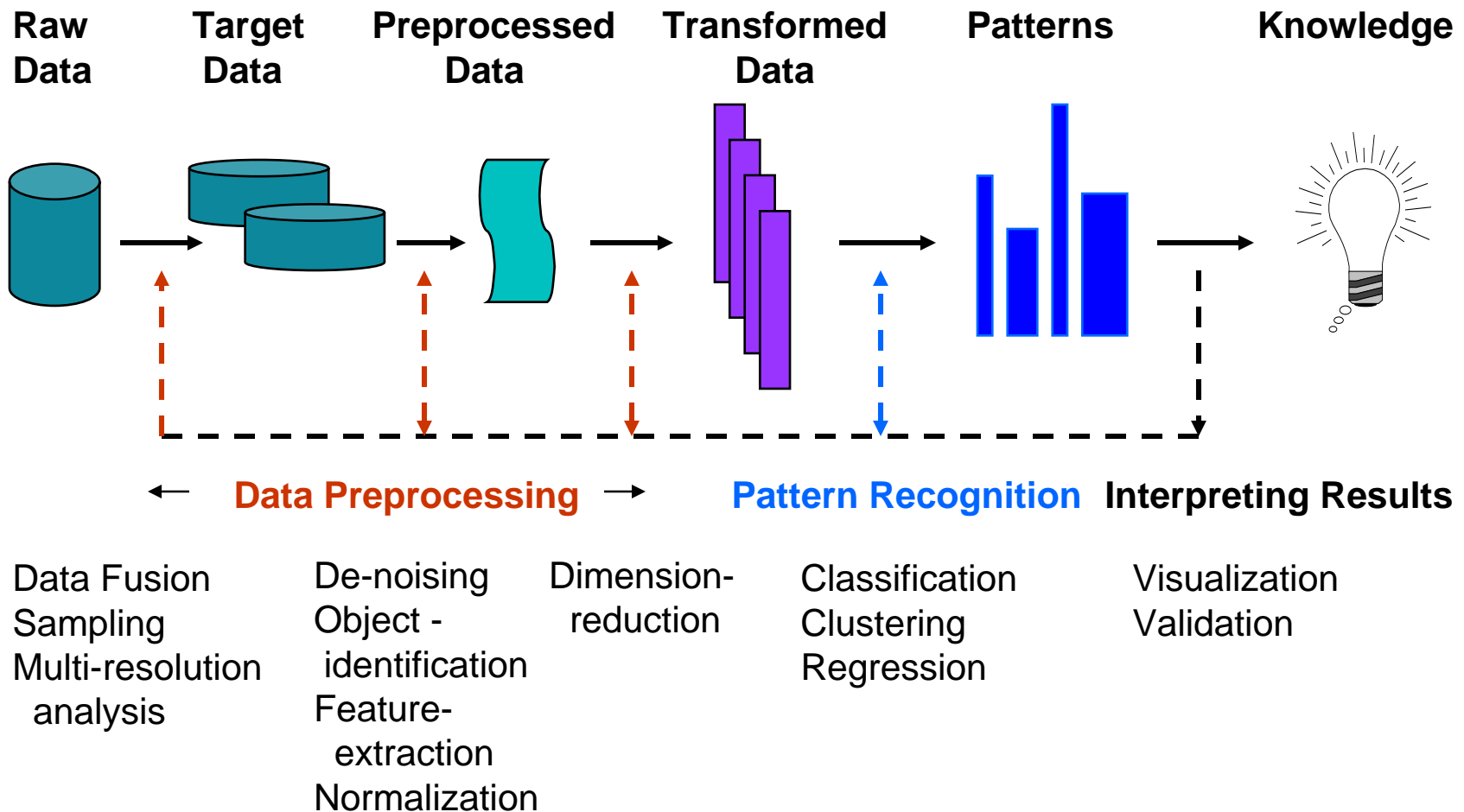*February 16, 2005*

http://www.llnl.gov/casc/sapphire

Sapphire: using data analysis techniques to address the data overload problem

- **Goal: analysis of science data from experiments, observations, and simulations**

- **Sapphire has a three-fold focus**
  - **analysis of data from practical problems using**
  - **modular, extensible software which incorporates**
  - **research in robust, accurate, scalable algorithms**

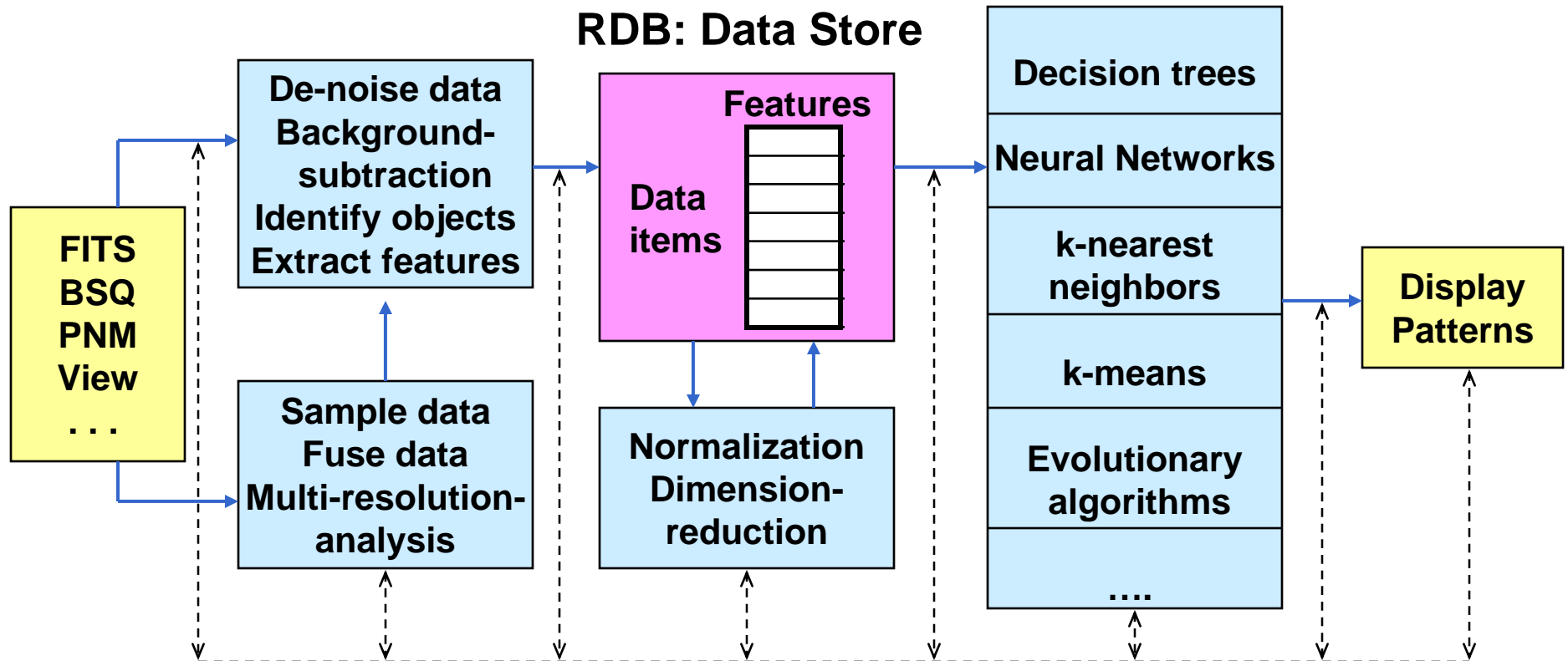- **Research areas – image processing, pattern recognition, information retrieval, motion tracking, …**

CASC

Large-scale data mining - from a Terabyte to a Megabyte

Raw Data	Target Data	Preprocessed Data	Transformed Data	Patterns	Knowledge

← **Data Preprocessing** →   **Pattern Recognition**   **Interpreting Results**

Data Fusion Sampling Multi-resolution analysis	De-noising Object - identification Feature- extraction Normalization	Dimension- reduction	Classification Clustering Regression	Visualization Validation

An iterative and interactive process

The Sapphire system architecture: flexible, portable, scalable

RDB: Data Store

FITS BSQ PNM View . . .

De-noise data Background-subtraction Identify objects Extract features

Sample data Fuse data Multi-resolution-analysis

Features

Data items

Normalization Dimension-reduction

Decision trees

Neural Networks

k-nearest neighbors

k-means

Evolutionary algorithms

….

Display Patterns

	Sapphire Software		Public Domain Software		Sapphire & Domain Software

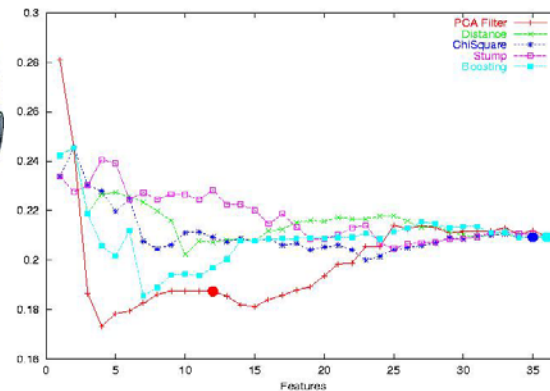**Components linked by Python**   <----> **User Input & Feedback**
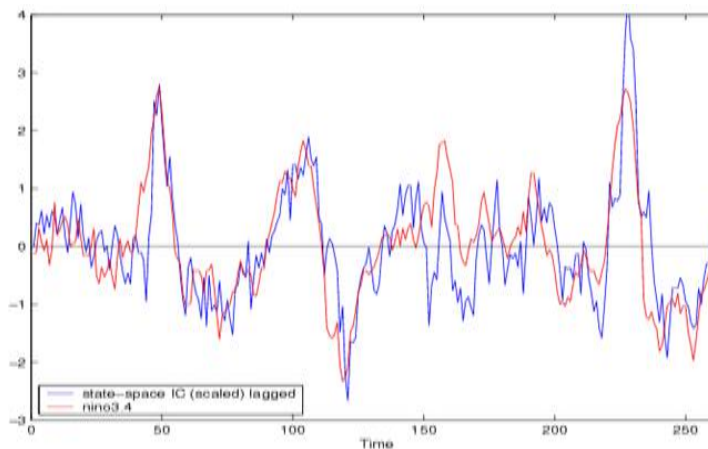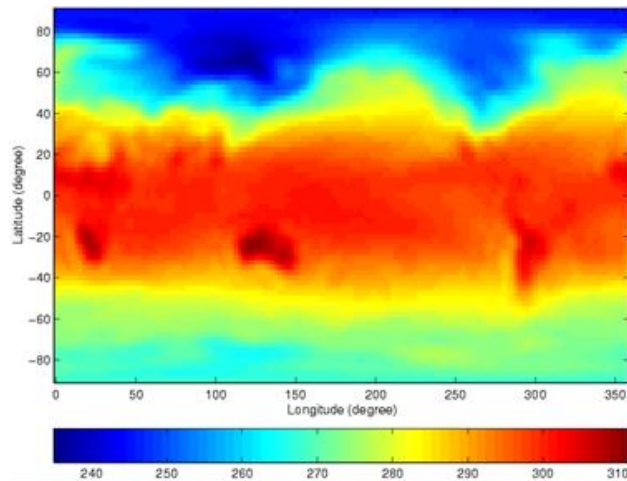
CASC

CK /Sapphire-4

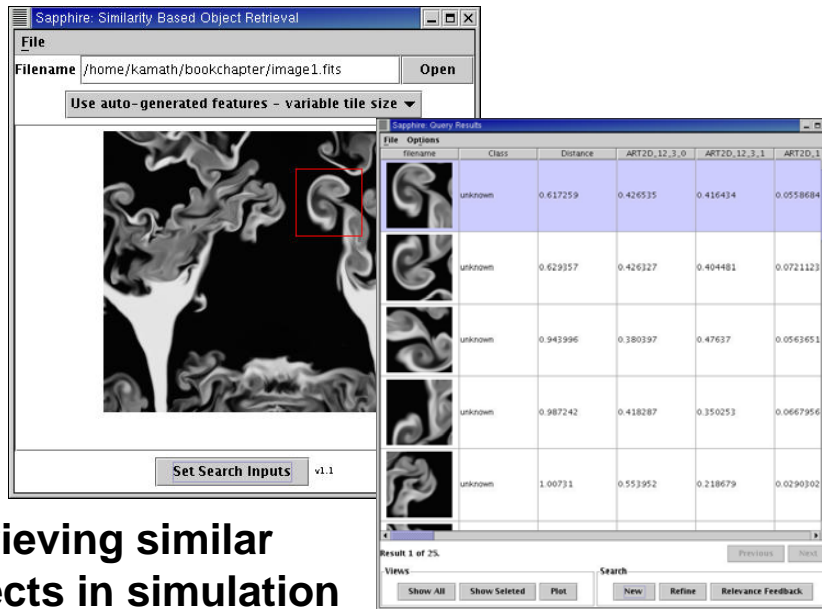# We have analyzed data from astronomy, climate simulations, experimental physics, …
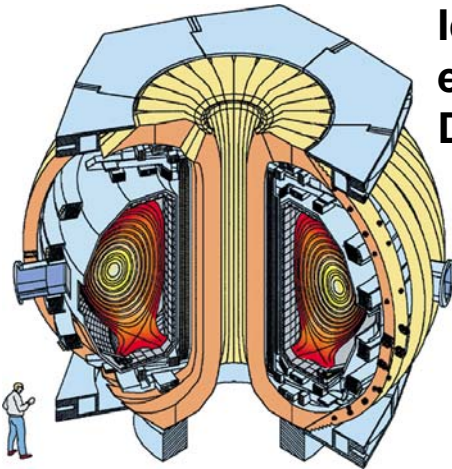
Identifying features connected to edge-harmonic oscillations in DIII-D Tokamak
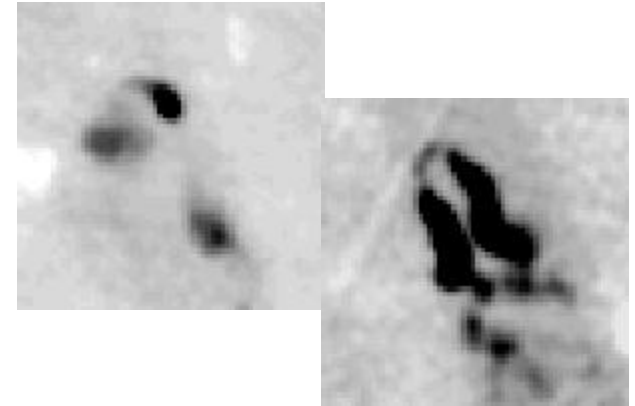
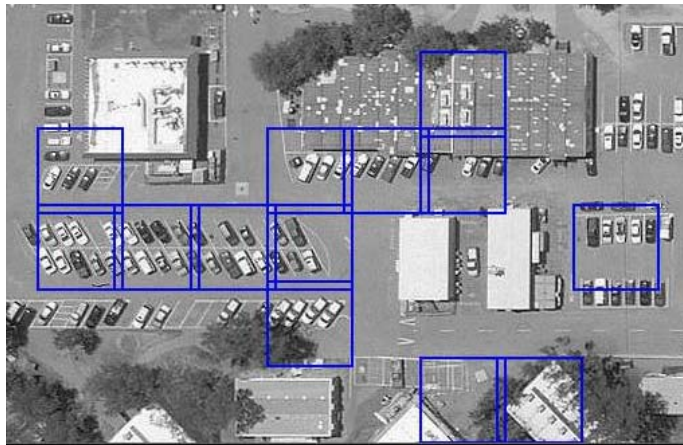**Classifying bent-double galaxies in the FIRST survey**

Separating signals in climate data

CASC

… remote sensing, and information retrieval

Finding human settlements

Satellite image by Space Imaging

Retrieving similar objects in simulation data ….

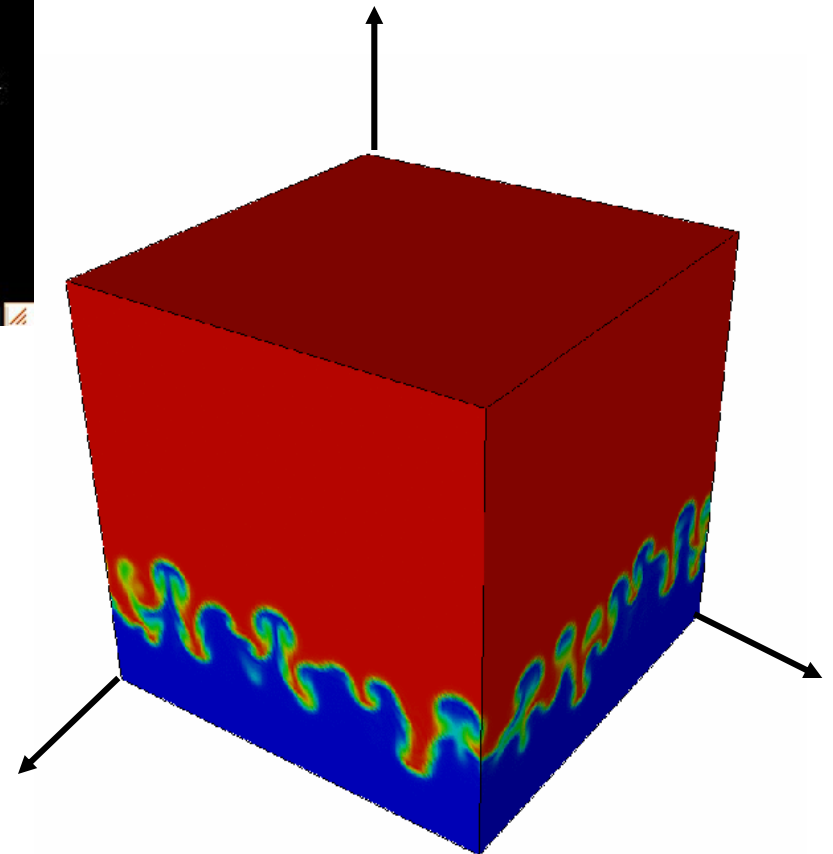**…. and remote sensed imagery**

CASC

Our current work includes 3-D fluid mix problems, experimental physics,…

Classification of Poincaré plots
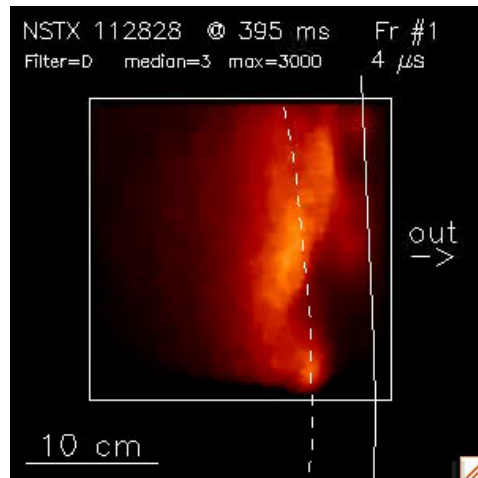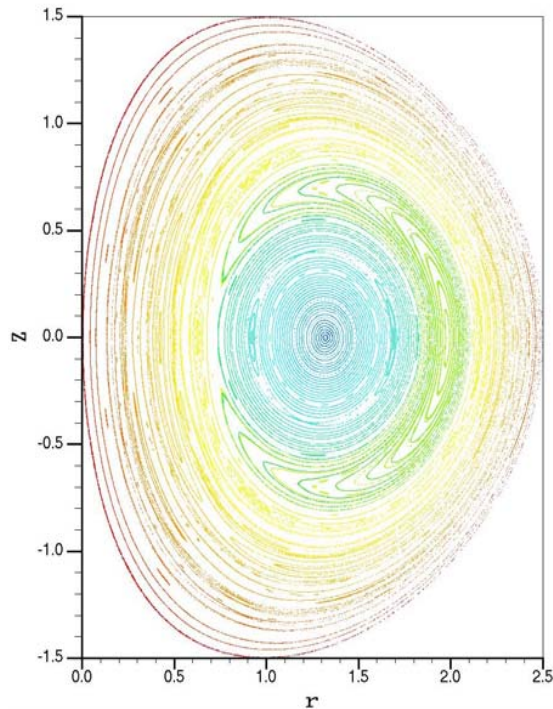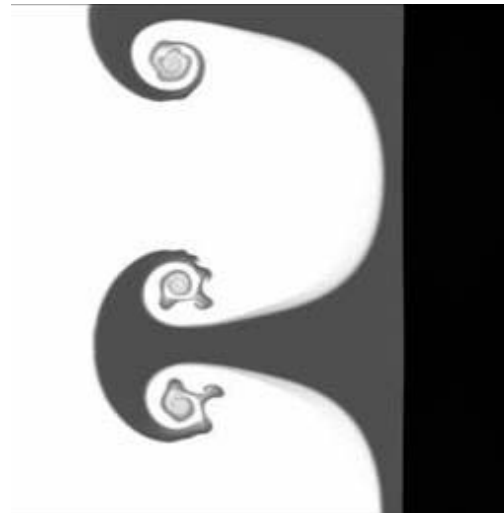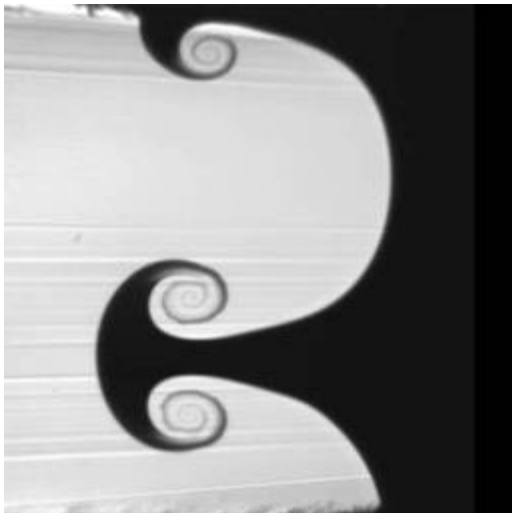
Tracking plasma

Analysis of 3D simulations

… detection and tracking of moving objects, and comparing simulations to experiments

Detecting and tracking moving objects in video

Comparing simulations to experiments

CASC

Analysis of Remotely-sensed Imagery

Application: finding human settlements in satellite imagery

Satellite image by Space Imaging

400x400 pixel image from the Nebraska region

Original satellite image by Space Imaging

Regions from which the training set was extracted

We use pattern recognition to build a model to identify the type of each pixel

Original image

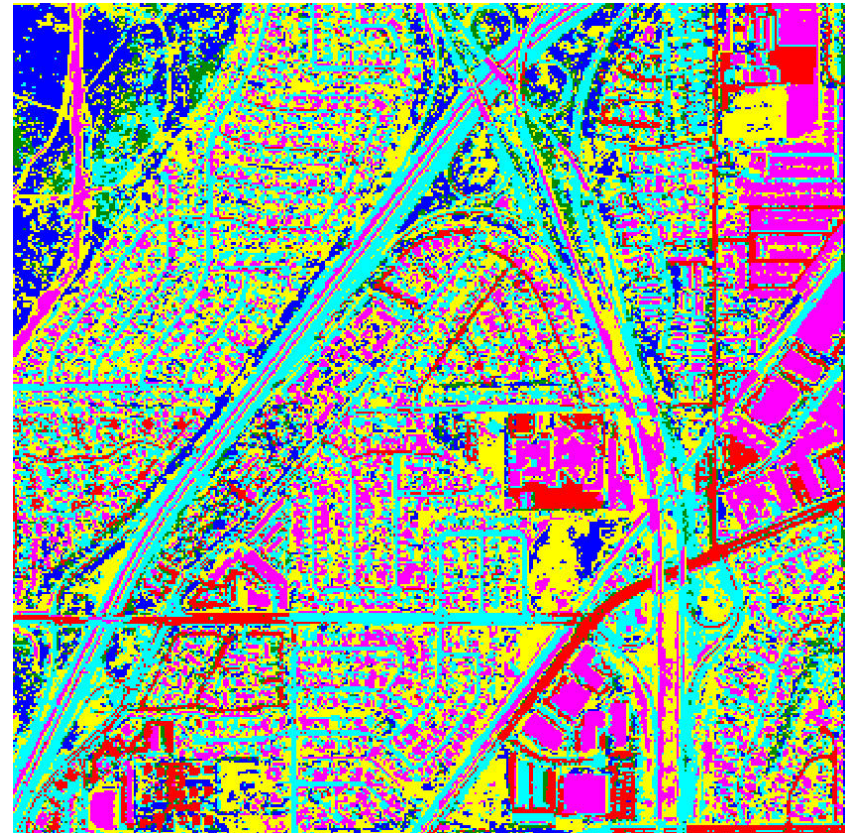**Pixels with types**

Satellite image by Space Imaging

The same model can also be applied to a different region

Original image from Northern Mexico

Satellite image by Space Imaging

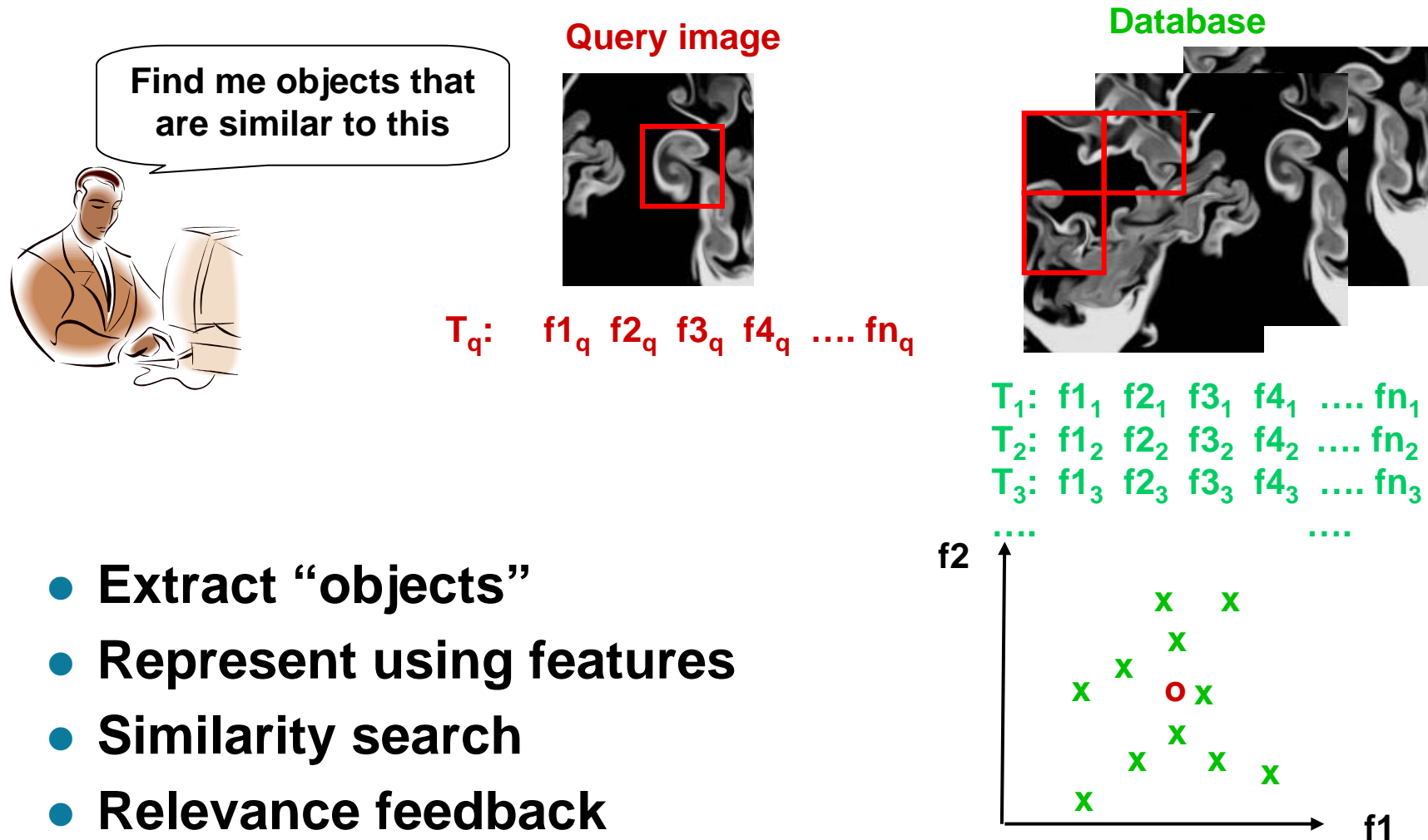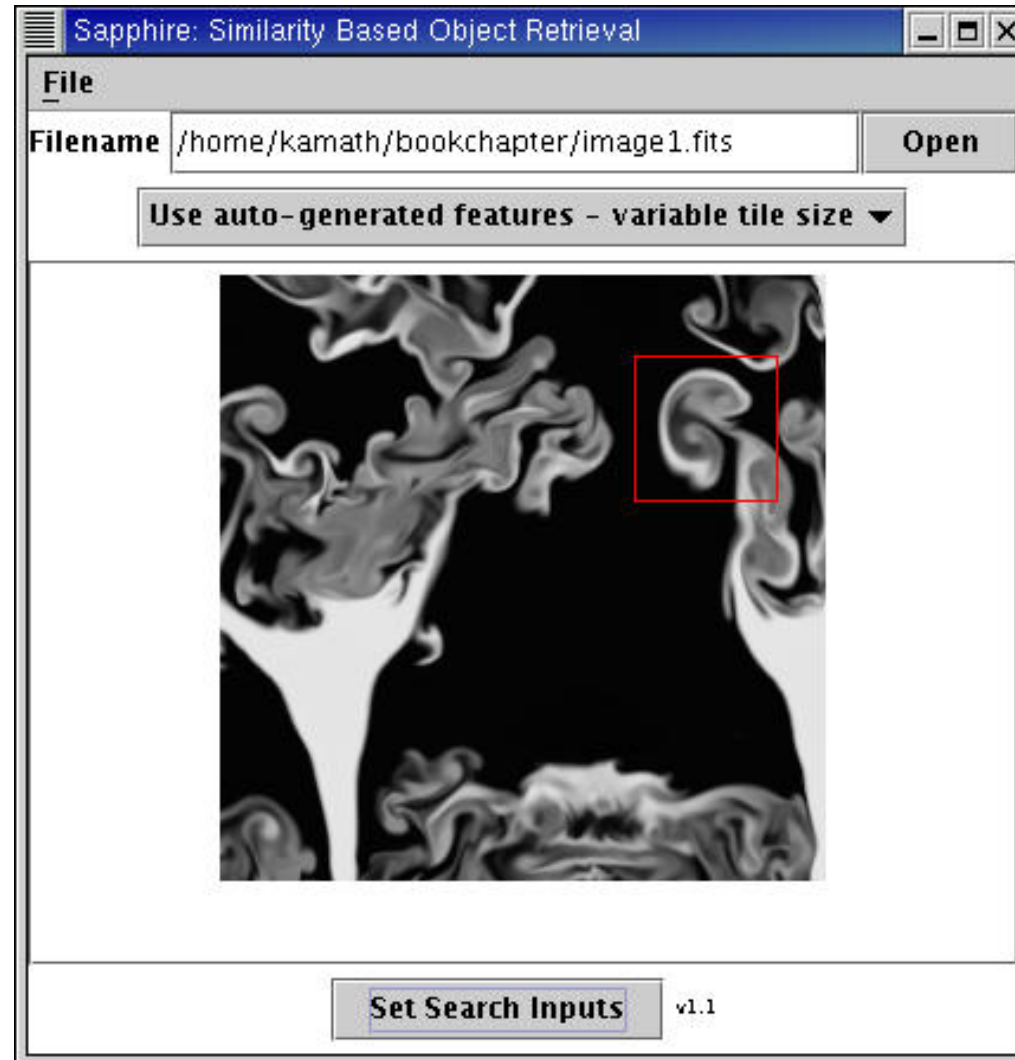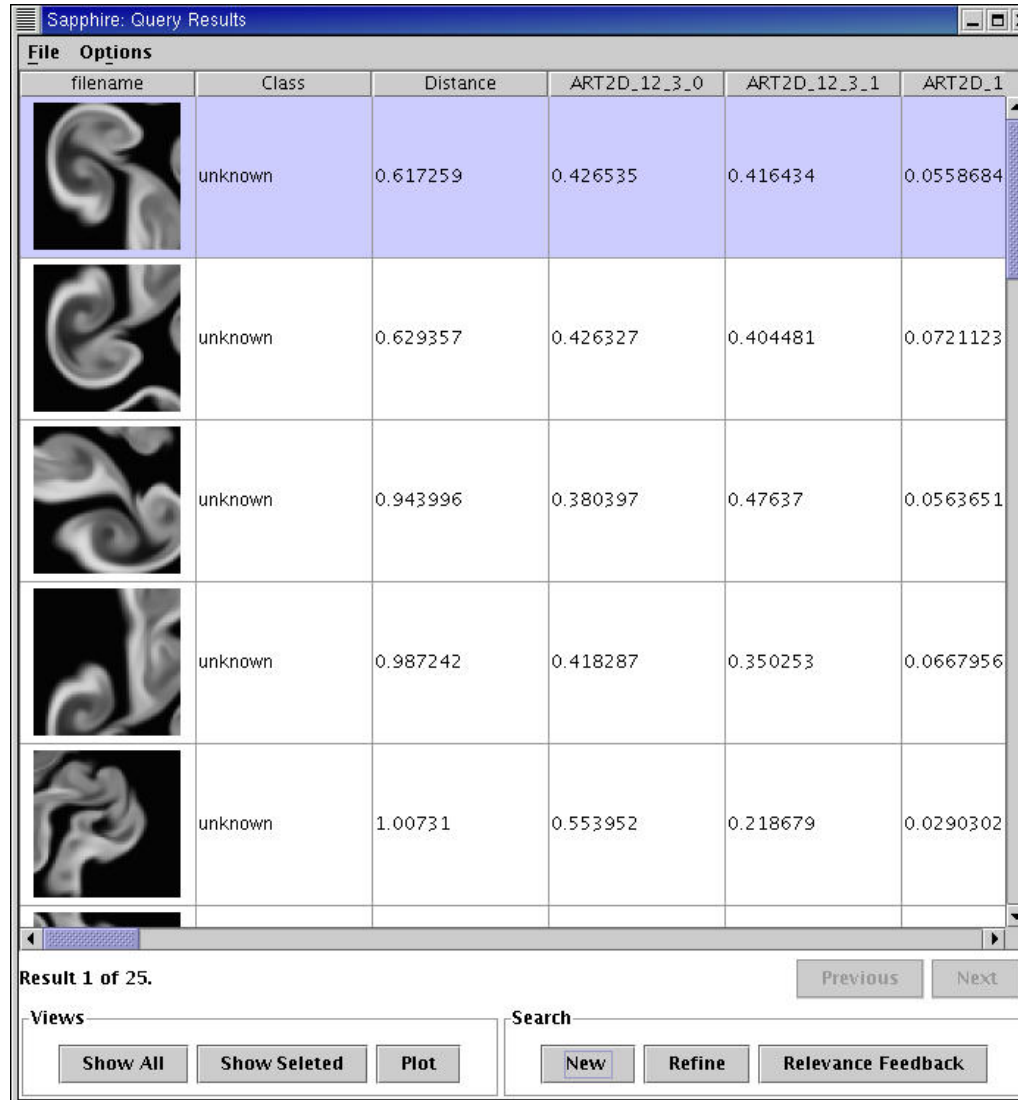**Pixels with types**

Finding Similar Objects in Image Databases

Given an "object", retrieve similar "objects" in a database

Find me objects that are similar to this

Query image

T_q: $f1_q$ $f2_q$ $f3_q$ $f4_q$ …. fn_q

Database

T_1: $f1_1$ $f2_1$ $f3_1$ $f4_1$ …. fn_1
T_2: $f1_2$ $f2_2$ $f3_2$ $f4_2$ …. fn_2
T_3: $f1_3$ $f2_3$ $f3_3$ $f4_3$ …. fn_3
….   ….

- **Extract "objects"**
- **Represent using features**
- **Similarity search**
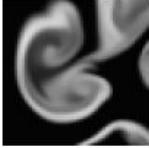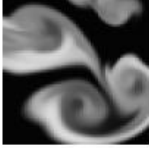- **Relevance feedback**

$f2$

$f1$

Our retrieval tool allows us to select a region of an image to query ....

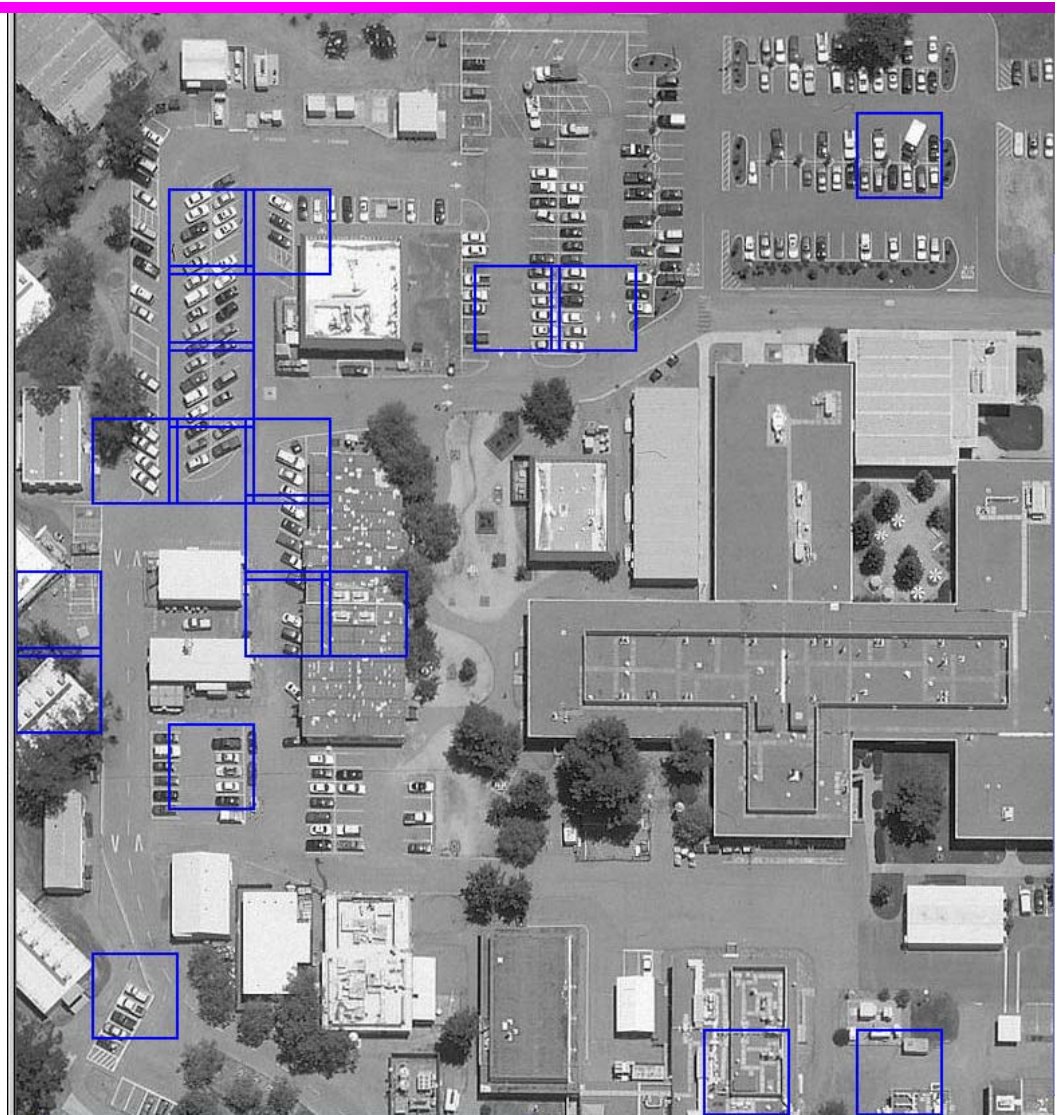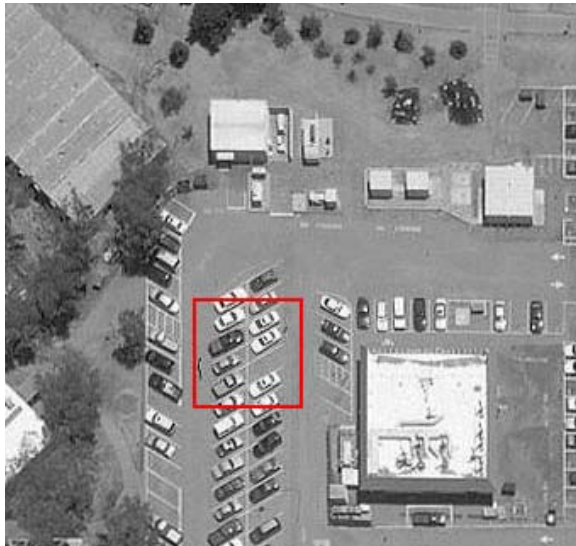# … and presents the results ordered by similarity

CASC

Our tool can be applied to data from simulations and observations

Detecting and Tracking Moving Objects in Video

Problem: track all vehicles that enter or leave the highlighted regions

Video is key to monitoring and surveillance

- **Can detect interactions between objects in the video**
  - **use to identify unusual events**
- **Several applications**
  - **movement of people, vehicles, ships, …**
  - **monitor traffic, detect pedestrians, lip reading, …**

Detection and Tracking In Video

➜ **Goal: identify unusual events in video**
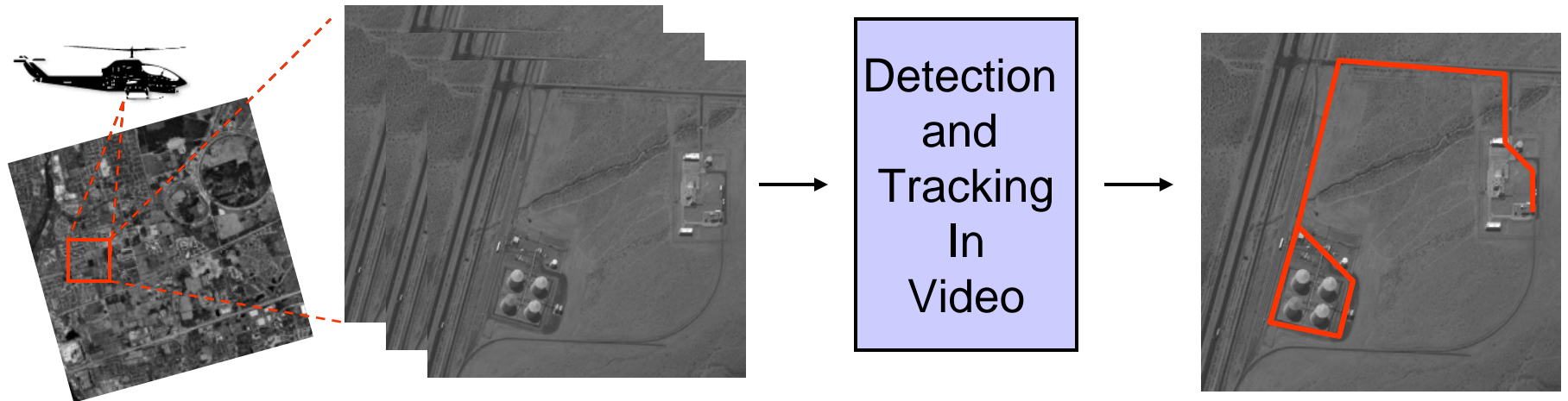
Our focus: low to moderate resolution video taken under less-than-ideal conditions

High resolution
~1 cm/pixel
Small FOV

Moderate resolution
~20 cm/pixel
Moderate FOV

Low resolution
~1 m/pixel
Large FOV

- **Characteristics of the data**
  - **a wide field of view (4K by 4K pixels)**
  - **low resolution (1 m/pixel)**
  - **low frame rate (1 frame/sec)**
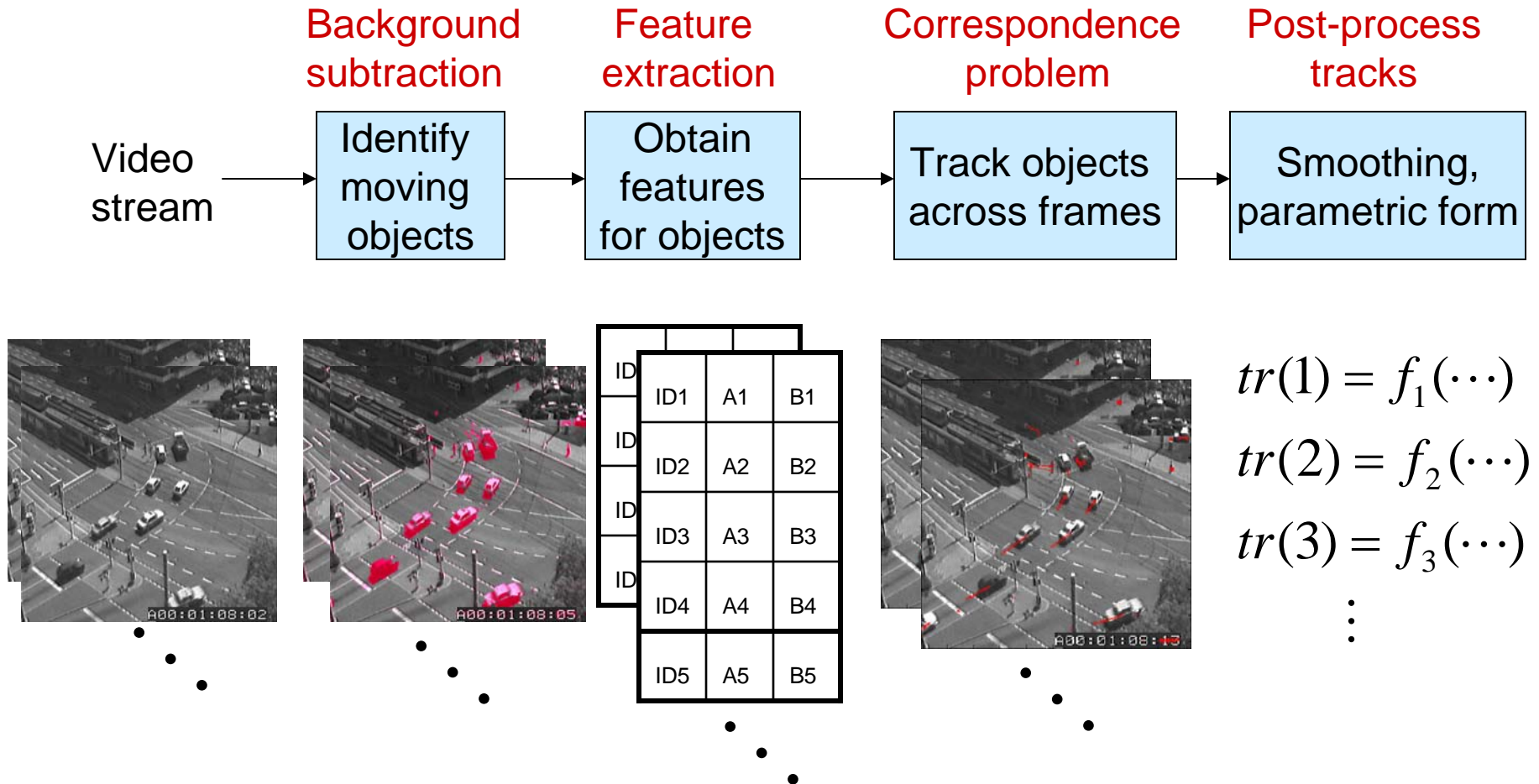
➔ **What is the resolution and frame rate needed to enable robust, real-time detection and tracking?**

CASC

Our approach analyses each frame, using the frames in a window around it

Video stream → | Background subtraction | Feature extraction | Correspondence problem | Post-process tracks |

| Identify moving objects | Obtain features for objects | Track objects across frames | Smoothing, parametric form |

ID1	A1	B1
ID2	A2	B2
ID3	A3	B3
ID4	A4	B4
ID5	A5	B5

$$tr(1) = f_1(\cdots)$$
$$tr(2) = f_2(\cdots)$$
$$tr(3) = f_3(\cdots)$$

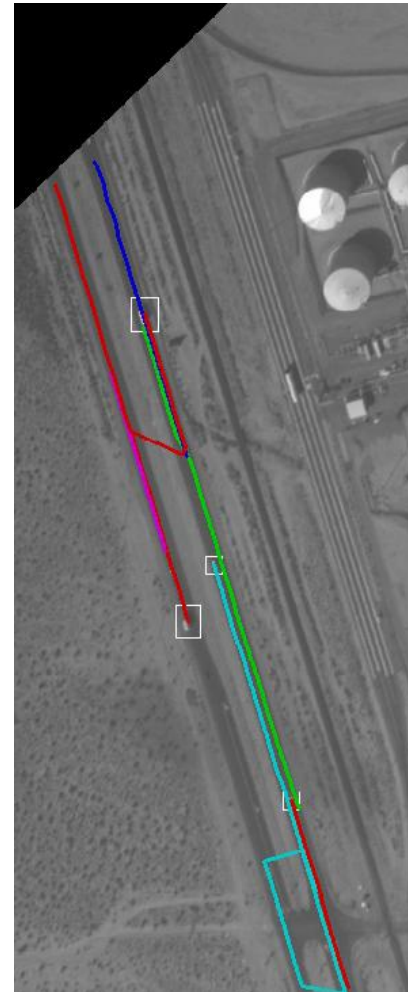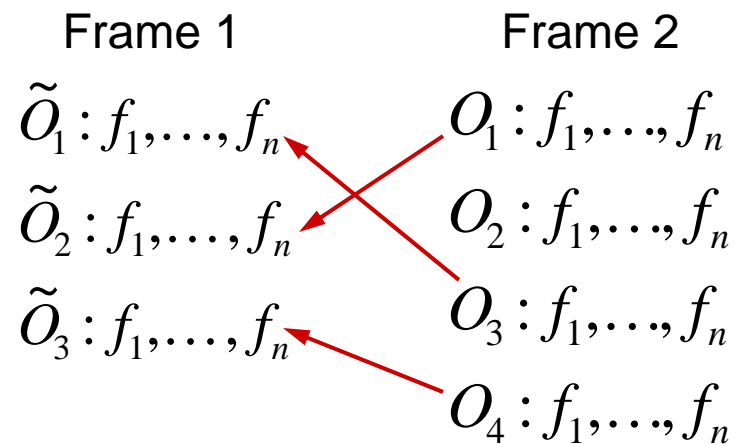# Our research led to a robust approach to detecting moving objects

Current best method

Our new approach

We found that complex techniques are necessary for effective tracking

Mapping objects across frames

Frame 1   Frame 2

$$\tilde{O}_1 : f_1,\ldots,f_n \qquad O_1 : f_1,\ldots,f_n$$

$$\tilde{O}_2 : f_1,\ldots,f_n \qquad O_2 : f_1,\ldots,f_n$$

$$\tilde{O}_3 : f_1,\ldots,f_n \qquad O_3 : f_1,\ldots,f_n$$

$$O_4 : f_1,\ldots,f_n$$

Simple tracker   Kalman filter

We are conducting research to make tracking more robust

Possible problems when a
car stops for a while

Problems in not identifying
objects correctly

A00:01:09:1-1

We would like our algorithms to be robust enough to easily handle difficult problems

Our tentative plans for DHS-funded work for CY05

- **Multi-media mining: combine text, audio, images, video, …**